I. Regression

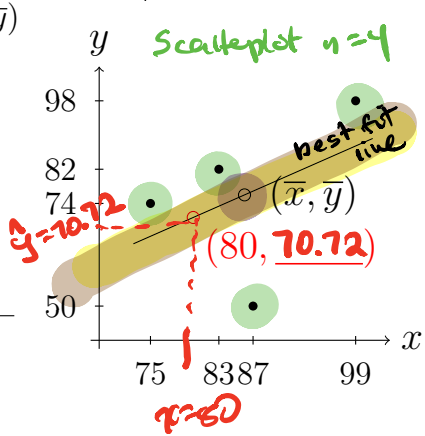
Scatterplots Correlation Regression

Regression: finding the best fit line

Given the 4 data below, find the regression line prediction $\hat{y}$ for $x = 80$.

a data point x	y	deviations $x-\bar{x}$	$y-\bar{y}$	square deviations $(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
75	74	$75-86=-11$	$74-76=-2$	121	4	22
83	82	$83-86=-3$	6	9	36	-18
87	50	$87-86=1$	-26	1	676	-26
99	98	$99-86=13$	22	169	484	286
sum:	344	304		300	1200	264
$\frac{\text{sum}}{n}$:	86 $=\bar{x}$	76 $=\bar{y}$				
$\frac{\text{sum}}{n-1}$:				100 $=s_x^2$ (variance)	400 $=s_y^2$	88 $=c$ covariance
				10 $=s_x$	20 $=s_y$	0.44 $=r$
						0.88 $=b_1$

$\frac{300}{3}=100$, $\sqrt{100}=10$, $4-1=3$, $\frac{88}{(10)(20)}=0.44$



Correlation $r = \frac{c}{s_x s_y}$

always between -1, 1

positive r means positive correlation

$b_0 = 0.32$ $b_1 = 0.88$

$\hat{y} = 0.32 + 0.88x$

Regression line simplified $\hat{y} = b_0 + b_1 x$

Regress line unsimplified $\frac{y-\bar{y}}{s_y} = r\frac{x-\bar{x}}{s_x}$

standard deviation

$y-\bar{y} = (rs_y/s_x)(x-\bar{x})$

$0.88 = rs_y/s_x$

Final answer $= \hat{y}(80)$ 70.72

$\hat{y} = 0.32 + 0.88x$
$= 0.32 + 0.88(80)$
$= 0.32 + 70.4$
$= 70.72$

$y - 76 = 0.88(x - 86)$
$y - 76 = 0.88x - 75.68$
$y = 0.88x + 0.32$